DOCUMENT RESUME

ED 123 667 CS 501 374

AUTHOR Lull, James T.; Cappella, Joseph N.
TITLE Slicing the Attitude Pie: A Validity and Reliability Study of a New Attitude Assessment Technique.
PUB DATE 76
NOTE 27p.; Paper presented at the Annual Meeting of the International Communication Association (Portland, Oregon, April 1976); Parts of text may be marginally legible due to print quality of original; Reproduced from best copy available

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Attitudes; *Attitude Tests; Educational Research; *Measurement Instruments; *Predictive Measurement; *Predictive Validity

ABSTRACT

This paper introduces a new measure of attitude, the "attitude pie," which attempts to separate true neutrality from degree of indifference while providing an evaluation score on the issue being judged. A large reliability-validity study involving ten attitude traits, four different methods, and over one hundred subjects at three different times was undertaken. Results from the study indicated that the reliabilities of both the salience and positivity scores of the attitude pies behave acceptably as compared to Semantic Differentials, Likerts, and Rank Importance procedures. Similarly the validity of the positivity scores is roughly comparable to that of the Semantic Differentials and Likert scales for evaluation. On the basis of the evidence from this study it was concluded that the attitude pie is at least as useful a tool in assessing group positivity and group indifference as either Semantic Differential or Likert methods. (TS)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED123667

Slicing the Attitude Pie: A Validity and Reliability Study of a New Attitude Assessment Technique

James T. Lull
Department of Speech
University of California,
Santa Barbara

and

Joseph N. Cappella
Department of Communication Arts
University of Wisconsin-Madison

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

James T. Lull

Joseph N. Cappella

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

BEST COPY AVAILABLE

Paper presented to the Information Systems Division of the International Communication Association Convention
Portland, Oregon
April 1976

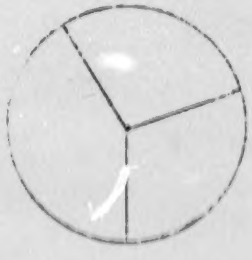
CS 501 374

The various techniques of attitude measurement must presume that the true, underlying attitudinal construct is being represented in the observed scores. However, most paper-and-pencil measures such as the widely used Likert and Semantic Differential procedures avoid treating the ambiguity associated with the so-called "neutral" response. When the subject checks the middle response on these instruments, the experimenter cannot be sure if the subject is completely indifferent or if the subject has judged the issue thoroughly and is equally favorable and unfavorable.

This paper introduces a new measure of attitude which purports to separate true neutrality from degree of indifference while providing an evaluation score (which we call "positivity") on the issue being judged. Thus the procedure provides a measure of affective favorability and unfavorability on the issue, as do the more standard paper and pencil techniques, as well as a measure of indifference which purports to separate those who are apathetic on the issue from those who are truly neutral. It is this latter measure which is the key point of difference between what we call "attitude pies" and the more typical linear scales exemplified by Likert and Semantic Differential instruments.

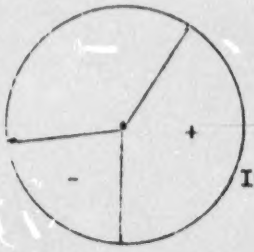
With the typical linear scales respondents are asked to indicate their feelings about attitude objects by simply marking a single space along a continuum. The bi-polar scale used with Semantic Differential and Likert scales is replaced by a circle or 1 3/8" "pie" with a ¼" dot placed in the center:

Subjects are told to consider the pie as representative of "all your feelings" about the attitude object or concept in question. They are given instructions to respond by drawing in three "slices" on the pie to represent the degree of positive, negative and indifferent components of their attitude toward the object. The slices drawn in, then, are thought to reflect relative amounts of positive and negative affective regard for the attitude object as well as the degree of indifference associated with the object. A typical pie response with roughly equal positive and negative affect, and with a good slice of indifference would appear as:



Positive, negative and degree of indifference scores are derived by employing a protractor to measure the degree of arc subtending each of the sectors. Since the sum of the angles in a circle must equal 360 degrees, only two angles need be measured to generate scores for the three attitudinal components. The scores that result show relative amounts of positive and negative feelings toward an object or concept as well as the overall intensity of these responses.

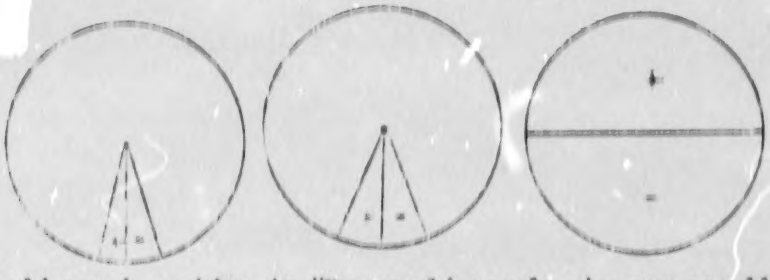
By measuring the angles of the various slices, four scores are derived. The positive score is simply the size of the angle representing that component of the respondent's favorability, the "plus" slice. Similarly, the negative score is the size of the angle of the "minus" slice and is presumed to be the subject's unfavorability on the issue. Positivity is the positive score minus the negative score and obviously it can range from -360 to +360. In the example immediately below the positive score (150) minus the negative score (80) produces a positivity score of 70. This yields a single evaluative score for each attitude object which is comparable to the scores generated by Likert and Semantic Differential scales.



+ = 150
- = 80
Positivity = +70
Indifference = 130

Indifference scores are derived by subtracting the sum of the positive and negative scores from 360.

The analysis of indifference scores allows the researcher to categorize positivity scores into classes or along a continuum. For example, in the three pies below the positivity is zero for each but the degree of indifference is decreasing from left to right. The question that every research who has used Likert or Semantic



Differential should now be asking is "How would people who are equally positive and negative but more or less indifferent respond on the standard scales?" Similarly are individuals who are slightly positive (or negative) and highly indifferent more likely to check the midpoint on a Semantic Differential or Likert than individuals with the same positivity score but with less indifference?

The reasons that responses to these questions do not immediately come to mind reflects the fundamental ambiguity of the midpoint on Semantic and Likert scales. Quite simply, we are arguing that to disambiguate the midpoint of the Semantic or Likert scales requires at least two pieces of information, one related to positivity of affect and one related to indifference or apathy (or its complement salience). In principle, the pies provide this information. Whether they do so in practice is the overriding empirical question.

A part of the empirical question is whether subjects will provide both indifference and positivity information on the pies. Earlier research with the pies showed that individuals can and do use the pies to show feelings of indifference and positive and negative affective reactions toward a variety of attitude objects (Lull and Arntson, 1974). In a sample of 41 persons who evaluated ten attitude objects nearly everyone (97.5%) reflected a component of "indifference" at least once. Ninety percent of the sample used the indifference component in combination with positive

and negative slices on at least one of the items. Subjects were also likely (70.7%) to mark a pie as either all positive or all negative on at least one of the ten objects. Thus people seem willing to use the range of responses which the pies provide.

In the following pages we wish to argue both the theoretic and practical advantages and disadvantages of the pies, describe a study designed to test the reliability and validity of the pies versus Likert-type and Semantic Differential competitors, and present data from that study which is both favorable and unfavorable to the pie instrument. Let us first take up the advantages and disadvantages of pie measures both in substantive applications and as a measurement tool.

THE PROS AND CONS OF PIES

There are at least three areas of current theory and research on attitudes that could be illuminated by employing pie-type measurement procedures. First in the area of interpersonal attraction and balance recent work has been careful to distinguish between pairs of people who are neutral toward one another and pairs who are indifferent to one another. As Hunter (1974, pp. 26-34) points out the various theories of how people group in sociometric networks (balance, positive balance, clustering, and his own dynamic invariante) have different predictions depending upon how "non-indifferent" links are treated. Of course indifferent links must be able to be separated from those which are not if theories are to be tested. The pies provide a convenient method for doing this. In a similar vein Cartwright and Harary (1972) have argued that the generalization of structural balance theories to networks "provides a means for representing more realistically the full range of evaluative relationships [among people] that are encountered empirically" (p. 498). In order to achieve this goal they introduce at a theoretical level the value, or measure, of a directed line as the ordered pair, p for positive, and n for negative, where p represents positive attraction to another and n disliking. Their triangle model which introduces a positivity score, p-n, an intensity score, p+n, and the condition that p+n is bounded by 1, is quite conveniently tested with the pie instrument. Balance theories and their extensions are cognizant of and in need of measurement procedures which the pies readily provide.

Secondly, in the attitude formation and change area there is some growing interest (especially by Lashbrook and his associates: Hylton and Lashbrook, 1972; Lashbrook and Bush, 1973; Lashbrook and Sullivan 1973; Sullivan and Lashbrook, 1974; Troyke and Lashbrook, 1973; W. Lashbrook and V. Lashbrook, 1975) in the behavior of audiences which are apathetic as opposed to those which are neutral. Apathetics and neutrals have been distinguished using the Fishbein and Raven (1962) attitude and belief scales, saliency scales, and expenditure of effort scales (as cited by W. Lashbrook and V. Lashbrook, 1975). The pie measure can provide the necessary classifications as a standard part of the information provided by an audience in their favorability ratings rather than as information over and above that provided by favorability scores.

Thirdly, the pie technique offers some intuitively satisfying information to lay consumers of research on public attitudes. In a study currently being carried out by one of the authors on audience recognition and favorability toward regional radio stations, the pies are providing both types of information simultaneously and in a form which is readily comprehensible to radio station managers. In particular the indifference score is readily associated with lack of recognition and the positivity score with favorability. Placing average pies for various demographic groupings of the audience side-by-side facilitates the comparison of subgroups without recourse to numerical (usually percentage in this case) summaries. While not directly pertinent to theory or theory-related research this dimension may serve as the strongest recommendation for the pies to the broader audience.

Despite the above described advantages, there are drawbacks to the pie instrumentation. First, the pies necessitate that subjects make judgments which are less familiar to them than judgments made on the more common 5-point Likert-type and 7-point Semantic Differential scales. While people often see for example, the "tax dollar" sliced in news descriptions of government spending and, hence, find it familiar enough to comprehend they seldom are asked to make judgments which slice their own attitude pies. This lack of familiarity will be likely to increase the pie's unreliability. Besides asking for unfamiliar types of judgments the pies require more complex judgments. The increased complexity can be seen in two ways: (1) Three types of information (although only 2 are independent) are required by each pie judgment--positive sentiment, negative sentiment, and indifference--as opposed to the one dimensional judgment of the Likert and Semantic Differential techniques. (2) The subjects express judgments on a circular continuum with no set number of categories. Discriminations by subjects can be as fine as they wish (although our protractors usually fix a lower limit of a 5 degree slice). While finer discriminations permit more subtle differences, they are also likely to increase unreliability. By standard criteria for measurement reliability (Torgerson, 195 , p.), the pies are likely to exhibit greater unreliability than the simpler and more familiar Likert and Semantic Differential instruments. While the degree of difference in reliability between pies and competitors is an empirical matter, we hope that the theoretic advantages of the pies are an acceptable trade off for a tolerable decrease in reliability.

In this section we have argued for the potential advantages of the pie instrumentation over its more accepted counterparts. Now we must subject our arguments to the critical scrutiny of empirical test. In particular a study of the reliability and internal validity of the pies was undertaken against the competitive techniques of Likert scaling and Semantic Differential scaling. These competitors were chosen because of the extent of their use by researchers and because a recent study (Jaccard, Weber and Lundmark, 1975) has shown that these two techniques are highly reliable over time (with test-retest correlations all above .80) and that they exhibit considerable convergent validity as compared to fancier Guttman and Thurstone scales. Our choice of Likert and Semantic Differential scales as baseline comparisons for the reliability and internal validity of the pies should bias all of our conclusions about the pies toward the conservative.

TESTING THE ATTITUDE PIES

The testing procedure was a variant of the basic multitrait-multimethod (MTMM) technique advanced by Campbell and Fiske (1959). The traits included 10 different attitude items. The methods included 7-point Likert-type scales, 7-point Semantic Differential scales, and attitude pies. The variation from straight MTMM procedures simply involved three waves of data gathered on the same subject samples rather than one. The availability of overtime data makes it possible to check the test-retest reliability of the methods and to check for learning effects (increased familiarity) with the pies.

TRAITS AND METHODS

Based on previous research by Lull and Arntson (1974) 10 attitude items were chosen as traits. They were chosen to maximize variance across traits on indifference and positivity scores. Thus "alka-seltzer," "radishes," and "polkas" were expected to be large indifference items. "Parents," "making love," and "inflation" were expected to be low on indifference and "summertime," "parties," "giving blood," and "handguns" were expected to be of intermediate value on indifference.

Each of the ten traits was evaluated by the subjects with at least two of the

methods at each point in time. At the third administration all three methods were used by all subjects. The instruction set for the attitude pies is included as an appendix to this paper. There is one pie to be sliced for each trait. For the Semantic Differential five scales for each trait were used. Four adjective antonym pairs were constant across the 10 traits: good-bad, positive-negative, meaningful-meaningless, and important-unimportant. Osgood, Suci, and Tannenbaum (1957, p.), report that all these pairs load cleanly on the evaluative factor. The latter two pairs were chosen because they were meant to tap a salience dimension for each trait. The supposition we made was that salience or importance would be inversely related to the indifference score of the pies. The fifth adjective antonym pair was chosen so that it loaded on the evaluative factor but was conceptually relevant to the trait being evaluated. Thus the pair "kind-cruel" is relevant to the trait "parents" but irrelevant to "polkas." "Worthless-valuable" is relevant to "polkas" but much less relevant to "parents." This technique is suggested by Osgood, Suci, and Tannenbaum (1957, p.).

Parallel to the Semantic Differential procedure five 7-point agree-disagree scales constituted the Likert-type items. The sentences which subjects were to disagree with or agree with (along the 7-point continuum) were constructed to be as semantically close to the Semantic Differential adjective pairs as possible. Thus two sentences of the five were designed to tap importance and meaningfulness. The other sentences sought to tap overall favorability-unfavorability for each trait, for example: "When I dance, I hate to polka" or "Summertime is pure joy." Furthermore, the 7 agree-disagree categories for Likerts were structured identically to the 7 categories of the Semantic Differential adjectives. Here again the similar structure of the instruments should favor the consistency between the Semantic Differential and Likerts rather than either of these and the pies. In order that the reader may convince himself or herself of the parallelism of the Semantic Differentials and Likerts, all the Likert sentences and the unique Semantic Differential adjective for each trait are reported in Appendix B.

ADMINISTRATION PROCEDURES

In the spring term of 1975 questionnaires were administered to an initial sample of slightly more than 100 undergraduate students in an introductory communication course at the University of Wisconsin. The students were distributed across 6 discussion sections of approximately 15-20 students per section. Three of the discussion sections were randomly assigned to one administrative group, Group A. The other three were assigned to Group B. The groups were approximately equal in size.

Group A received three administrations of the attitude questionnaire separated by one month delays. The first questionnaire asked Group A subjects to (1) to rank the attitude items in terms of importance, (2) to evaluate the items on pie scales, and (3) to evaluate the items on the previously described Semantic Differential scales. The second administration of the questionnaire was identical to the first. The third questionnaire was identical to the first two except that the previously described Likert scales were added.

Group B received the same treatment as Group A except they received the Likert scales and not the Semantic Differential scales at times 1 and 2. As with Group A the scale which did not appear at times 1 and 2 did appear at time 3. Table 1 summarizes the administration procedures and associated sample sizes.

Table 1 Here

Although some subject attrition occurred between times, it was not appreciable. Between times 1 and 2 Group A has a sample size of 51 and Group B of 53; for times 2 and 3 sample size is 49 for Group A and 46 for Group B; between times 1 and 3 there are 49 subjects with data in Group A and 46 in Group B. Across all 3 times there are 46 subjects in Group A and 43 in Group B.

SUMMARIZING THE DATA

In order to treat the data in MTMM terms, various indices were constructed. The rank order data remained the same with highest numbers indicating lowest rank. The pies were transformed into a positivity score (range -360 to +360) and on indifference score (range 0 to 360). The three purely evaluative items are the Semantic Differential were summed to yield an evaluative score (SDE) (range 3 to 21). The two other items were also summed to yield a salience score (SDS) (range 2 to 14). Similar indices were constructed for the Likert items (LE and LS) with the same ranges as the SDE and SDS indices.

Consider first the evaluation indices. Group A has 10 traits and 2 methods (SDE and positivity) at time 1, 10 traits and 2 methods at time 2, and 10 traits and 3 methods (LE is added) at time 3. Group B has the same number of traits and methods at the three times but the Likert method switches roles with the Semantic Differential method. Thus, there are 3 MTMM's to study (one for each time) for each group.

Next consider the salience and indifference indices. Group A has 10 traits and 3 methods (indifference, ranks, and SDS) at time 1, 10 traits and 3 methods at time 2, and 10 traits and 4 methods (LS is added) at time 3. Group B has the same number of traits and methods with the Likerts and Semantic Differentials switching roles. Thus, there are 3 MTMM's to analyze (one for each time) for each group.

Obviously this tremendous volume of data must be reduced to a manageable form and at the same time reveal the subtleties of the pie's performance relative to that of the Likerts and Semantic Differentials. We turn to this task in the next section.

RELIABILITY OF THE PIES

There are relatively few techniques available to the researcher for the purposes of estimating the reliability and validity of proposed measures. The reliability of a technique can be assessed either through its internal consistency (assuming parallel tests) or its consistency overtime (test-retest correlation). The validity of a measurement procedure can be evaluated in terms of its internal validity (correlation with another procedure presumably measuring the same trait) or predictive validity (correlation with an effect which is theoretically caused by the variable as measured by the technique under question). We restrict our evaluation of attitude pies to their over-time reliability and to their internal validity.

Our goal in the remainder of this paper is two-fold: (1) to present data on the reliability and validity of the pies and (2) to assess the relative merits of different techniques for determining reliability and validity of the pies. In order to accomplish both goals we shall very carefully describe the assumptions of each technique used and evaluate those assumptions relative to our own data. The description of assumptions will be facilitated by employing path diagrams (Heise, 1968; Duncan, 1971) whenever possible.

TEST-RETEST RELIABILITY

Perhaps the most time honored method of assessing over-time reliability is the straightforward test-retest correlation. Figure 1 represents the case of a single variable measured at two points in time. X* is the true value of the variable and X its observed counterpart. e(1) and e(2) are random errors of measurement at each of the two time points. u(2) represents variables which affect or disturb X during the time interval which has elapsed (Heise, 1971, p. 352). As can be seen from the path diagram, this measurement model assumes that the measurement errors are

---

Figure 1 Here

---

uncorrelated with true scores, that disturbances are uncorrelated with exogenous variables, and that the reliability p is stable over time (Heise, 1971, pp. 352-53). As can be easily seen from the diagram $r_{12} = p^2 p_{21}$ where r_{12} is the true correlation between X(1) and X(2). This equation cannot be solved unless some assumptions about p_{21} are made. Since p_{21} is a measure of the stability of the true scores over time, it is often assumed that $p_{21} = 1$ implying perfect stability. Under these assumptions the reliability p^2 is just r_{12}.

If we were to view the data on each trait as if it were gathered independently over the panels time-1 and time-2, and time-2 and time-3, then by the model of Figure 1 the reliabilities would just be the test-retest correlations. These are reported in Table 2.

Table 2

With the sample sizes as reported in Table 1, all the test-retest correlations reported in Table 2 are statistically significant at $p < .025$ at least. Even though all the reliabilities (except rankings for group B) show more consistency between times 2 and 3 than between times 1 and 2, none of these increases is statistically significant by the z-test (Hays, 1963, p. 532). These small learning effects are not

appreciable. Thus increased familiarity with the pies does not produce appreciable increases in test-retest consistency.

When the test-retest correlations for the pies are compared to those for the Semantic Differentials, Likerts, or Ranks, the pies invariably perform less consistently on both the salience and evaluation dimensions. However, although the differences in reliability favor the more standard measures none of the differences achieve statistical significance. The lack of statistically significant differences is an important matter since we must be concerned not with the absolute reliabilities of the pies but with their reliabilities as compared to other standard techniques under the same testing procedures. The absolute reliabilities of the pies are not as high as most researchers would like (the .7 to .9 range being more acceptable) but by the same token the reliability of the Semantic Differentials and Likerts is below that reported in the Jaccard, Weber, and Lundmark (1975) study.

Part of the reason for this depreciation in consistency over time can be attributed to the longer time delays in administering the retest questionnaires (1 month between administration) then is often allowed in research settings. With shorter time delays an increase in the unreliability of any scale is to be expected. A second reason for low test-retest correlations is that the test-retest correlation version of reliability assumes that the true scores are perfectly stable across time. With some of the traits employed, this assumed stability is highly unrealistic, especially with the long time delays between measurements. For example, "Inflation," was a national issue during the Winter and Spring of 1975 as the country slowly moved from a very unstable economic situation to an increasingly stabilized one. Similarly, "parents," "summertime" "making love," and "parties" could be dramatically affected by the information and experiences of a two month period. The point is that reliability should not be confounded with stability. A measure of reliability which separates out stability is needed. We consider such a measure next.

RELIABILITY INDEPENDENT OF STABILITY

In the diagram of Figure 1 we saw that in order to treat a test-retest correlation as a reliability it was necessary to assume that the stability between true scores at times 1 and 2, p_{21}, was perfect, or equal to 1. In other words, the situation is underidentified without further information. As Heise (1971) points out, if three waves of data are present the underidentified model of Figure 1 becomes identified and "a new measure of reliability based on test-retest data but free of temporal change effects" (p. 355) can be developed. Figure 2 describes the 3-wave situation and the associate assumptions. The assumptions for this case are quite

---

Figure 2 Here

---

similar to the case of Figure 1. In particular note that (1) the stability of the true scores p_{21} and p_{32} is not assumed to be the same between panels, (2) the errors of measurement are mutually uncorrelated so that test-retest sensitization is assumed negligible, and (3) the reliability, p^2, is assumed constant across time intervals.

With these assumptions it is an easy matter to show that a measure of the reliability which is independent of temporal change (Heise, 1971, pp. 354-55) is given by $p^2 = r_{12}r_{23}/r_{13}$. The advantage of this measure is that it does not assume anything about the degree or sign of the stability of the true scores between each of the 3 waves of data. It does, of course, assume that the rate of stability or instability is constant. On the other hand, this measure has the disadvantage of allowing peculiar values for reliability. The p^2 value above can be greater than 1 or can be negative. In both cases, the parameter value is uninterpretable and cannot

be used. Heise does not comment on this matter.

A model similar to Heise's was developed by Wiley and Wiley (1971). The Wiley and Wiley approach does not necessitate the assumption that the reliability is constant across time intervals and does not necessitate the use of standardized data. It does however assume that error variance is constant across time. Since the Heise reliability can be shown to be identical to the Wiley and Wiley reliability at time 2, then there is no advantage to the Heise scheme. In the tables which follow only the Wiley and Wiley indices of reliability will be reported. The reader should keep in mind the identification between the reliability at time 2 and Heise's measure. Wiley and Wiley also insist upon the use of unstandardized measures. We will follow suit allowing the variance inherent in each measurement method to influence the measures of reliability for each technique.

In Tables 3 and 4 Wiley and Wiley's measures of reliability (1971, p. 368) averaged across the 10 traits are reported for both groups at the 3 time periods. Table 3 reports the reliabilities of the evaluation measures and Table 4 reports reliabilities for the salience measures.

Tables 3 and 4

The positivity and indifference scores using the Wiley and Wiley indices for reliability are still somewhat smaller than the reliabilities of competitive scales within a given group and within a given time frame. However, the differences are not significant, averaging less than .1.

When the reliabilities from the 3-wave constant error variance model are compared to reliabilities from test-retest correlations (Table 2), three observations should be noted (1) overall the reliabilities on all scales are greater in the former case. (2) The pie reliabilities are less different from competitive reliabilities in the Wiley and Wiley model than in the test-retest model. (3) There is a non-significant but consistent trend toward greater reliability for the evaluation scales in both the Wiley and Wiley and the test-retest cases. While there is no way of testing which of the two models better fits the data (since both are exactly identified), we feel that there are obvious reasons to favor the Wiley and Wiley constant error variance model over the earlier one. When one considers the kind of traits employed in this study and the time logs between measurements then (to assume that true scores are perfectly stable across time) seems a mere convenience at best. If the Wiley and Wiley indices of reliability are to be preferred, then it is only fair to conclude that the pies both in the evaluative and salience dimensions are not appreciably less reliable than standard Semantic Differential, Likert, and Rank-order methods of assessing favorability and importance.

CORRELATED ERROR MODEL

Not being satisfied with the pies as "not appreciably less reliable," we sought to exploit other negative characteristics of Semantic Differential, Likert and Rank-order scales. In particular, the simplicity of judgments required by these procedures is both a boon and bane. On the positive side, simplicity of judgment helps to insure reliability or consistency. On the negative side, simplicity of judgment may enhance the liklihood of test-retest sensitization whereas the more complex judgments required by the pies may be less susceptible to sensitization. If this reasoning is accurate, then by including curved arrows (indicating correlation) among the errors of measurement e(1), e(2), and e(3), in Figure 2 and solving for the reliability, then test-retest sensitization might be accounted for.

Unfortunately, such a modified model is severely underidentified as Heise (1971, p. 358) notes. Fortunately, another set of Wileys (Wiley and Wiley, 1974) solved the underidentification for any 3-wave equal time lag data set by postulating a first-order autoregressive scheme among the errors. Unfortunately, the reliability indices for the correlated error model do not produce interpretable results for the data set of this study.

While this lack of interpretable reliability values is frustrating and incomprehensible to us, it is not totally unexpected since as Wiley and Wiley (1974, p. 185) point out: "The correlated error model does not yield admissable parameter estimates for all sets of data." Using the correlated error model algorithum reported by Wiley and Wiley (1974, p. 181) does not yield a single reliability estimate which is less than 1. All estimates fall in the range 1.0 to 2.2. Obviously, there are some conditions under which the correlated error model does not hold. Unfortunately, Wiley and Wiley have not specified when uninterpretable results are to be expected. Whatever these conditions are, this study satisfies them perfectly. Thus we must be satisfied with the reliabilities from the constant error variance case and live with the confound of sensitization.

VALIDITY OF THE PIES

Attitude pies offer two types of information: evaluation and salience. In order to establish the validity of the pies, validity of both the positivity scores and the evaluation scores must be established. This section will present data on the validity of the pies primarily through the multi-trait - multimethod technique (Campbell and Fiske, 1959).

MTMM MATRICES FOR EVALUATION SCORES

Referring back to Table 1, it can be seen that there are 6 potential MTMM matrice to be displayed: 3 for Group A and 3 for Group B. Rather than displaying the 1430 correlations from these six matrices we shall merely present the validity diagonals averaged across the appropriate groups and times. Figure 3 summarizes the structure of the averaged MTMM.

Figure 3 Here

As the figure indicates not all the monomethod-heterotrait blocks and heteromethod-heterotrait blocks are based on the same number of observations due to the asymmetry of the questionnaire administrations overtime.

The single most important portion of this matrix is the validity diagonal of each heteromethod block. The averaged diagonals for each heteromethod block of Figure 3 are presented in Table 5. The correlations reported are averaged across

---

Table 5 Here

---

groups and times as indicated in Figure 3. Campbell and Fiske's first criterion for validity (1959, p. 82), namely that "entries in the validity diagonal should be significantly different from zero and sufficiently large to encourage further examination of validity "is clearly met for 9 of the 10 traits in table 5. Only "inflation" does not exhibit the desired validity. This anomalous result is rather

befuddling and inexplicable to us. Aside from this single trait Campbell and Fiske's first and most important criterion is easily met by the positivity scores. As with the reliability measures previously, the scales competing with the pies tend to outperform them. The validity correlations between SDE and LE are stronger than between either of these scales and the pies for all traits except "love." Interestingly the highest intercorrelation of SDE and LE scales occur for the traits "guns," "blood" "polkas," and "radishes" which over the entire sample are ranked in importance 6th, 8th, 7th, and 9th respectively. For certain items which have been ranked in the upper 5 in importance, namely "parents," (#1) "love," (#3) and "parties," (#4), correlations between pies and SDE, and pies and LE are more similar to correlations between SDE and LE scales. Thus, in this data set when the pies fail to exhibit the same degree of convergent validity as Semantic Differential or Likert methods, it can often be explained in terms of the lower salience of the trait to the subjects. With respect to convergent validity then the pies perform about as well as that (that is, within .1 on the average) or better than Semantic Differentials and Likerts for 3 of the 5 most salient traits and 3 of the 5 least salient traits. The pies are clearly outperformed on "summer" (#2), "inflation" (#5), "guns," (#8), and "radishes" (#9).

MTMM MATRICES FOR SALIENCE SCORES

As with the evaluation scores, there are 6 MTMM matrices to be considered. In this case, however, averaging over Groups A and B for times 1, 2, and 3 produces a 10 trait by 4 method (pies, SDS, LS, and Ranks) matrix. As before we will apt not to present the averaged MTMM of 780 correlations but offer the six validity diagonals for each of the ten traits in Table 6. These are enough. Of all 6 validity diagonals

---

Table 6 Here

---

only the Semantic Differential and Likert methods show any pattern even vaguely resembling convergent validity. The higher correlations are somewhat misleading however since the "meaningful-meaningless" and "important-unimportant" adjective pairs which formed the SDS scale and were the basis for the LS agree-disagree statements were chosen because they loaded on the evaluative dimension in Osgood, Suci, and Tannebaum's 1957 volume. Thus there is some question as to whether these items can be said to tap salience primarily. In fact when SDS and LS indices are correlated with rank importance they cannot be said to satisfy Campbell and Fiske's first criterion for convergent validity cited above.

Of course the behavior of pie salience scores is not very encouraging either. The pies do not correlate significantly with questions on importance (SDS and LS) and do not correlate significantly with rank importance. Despite the inability of any of these measures to conclusively show convergent validity on salience that does not exempt the pies from doubt concerning the validity of the indifference score as a measure of salience.

Some of this doubt can be disspelled by computing mean salience scores for each trait on pie, ranks, and SDS for Group A across each of the 3 times and mean salience scores for each trait on pies, ranks, and LS for Group B across each of the 3 times. By ranking the traits on the basis of their mean salience as determined by each of the three methods, then Spearman rank correlations can be calculated. The mean salience scores for Groups A and B are presented with the corresponding rankings in Table 7. The Spearman r for pies and ranks in Group A is .964 and for SDS and ranks is .879. The Spearman r for pies and ranks in Group B is .916 and for ranks and LS

Table 7 Here

---

is .867. Thus, although we have no evidence that the salience score on pies can successfully discriminate an apathetic from a committed individual on an issue, there is some evidence that groups can be discriminated in terms of their apathy or commitment on an issue.

CONCLUSIONS AND DIRECTIONS

This paper has introduced a measurement tool for attitudes which attempts to separate individuals who are neutral on an issue from individuals who are apathetic. The tool seeks to accomplish this goal by simultaneously requiring individuals to provide information on their favorability toward an issue and their relative intensity of positive and negative affect toward the issue. In order to establish this tool as an effective instrument for research its reliability and validity in both the positivity and indifference components need to be established. A large reliability-validity study involving 10 attitude traits, 4 different methods, and over 100 subjects at 3 points in time was undertaken.

Results from our study indicate that the reliabilities of both the salience and positivity scores of the pies behave acceptably as compared to Semantic Differentials, Likerts, and Rank Importance procedures. Similarly the validity of the positivity scores is roughly comparable to that of the Semantic Differential and Likert scales for evaluation. However the SDE and LE scales do show some validity success beyond that of the positivity scores for certain traits. Unfortunately, validity data on the salience scores for the pies was very unfavorable although this evidence was mitigated by a failure to establish any mono-trait - heteromethod correlations as an established baseline for assessing convergent validity on the salience dimension. When we considered the validity of pie measures of salience compared to rank importance for group means rather than individual scores, the indifference scores behaved admirably.

On the basis of the evidence from this study we can safely conclude that the attitude pie is at least useful a tool in assessing group positivity and group indifference as either Semantic Differential or Likert methods. When the mean scores of groups are not the mode of analysis but individual scores are of prime concern, then the pies should provide estimates of positivity which are only a little less reliable and valid than Semantic Differential and Likert scales. However, the researcher should be hesitant to employ the indifference score when individual data are the prime requirement.

It is our hope that attitude pie will be adopted in the day to day work of researchers so that further data on predictive validity and reliability can be acquired. We feel and have tried to argue that there are some advantages to the pies beyond those afforded by standard linear measures. Obviously, some difficulties exist with pie measures and these must either be rectified by employing different instruction sets or must be more definitively tested so that the doubts about the indifference score may be either confirmed or laid to rest.

TABLE I

Administration procedures indicating scales given to each group at each point in time along with sample size

	TIME-1	TIME-2	TIME-3
Group A (Size)	56	53	50
Rank	Yes	Yes	Yes
Pies	Yes	Yes	Yes
Differential	Yes	Yes	Yes
Likerts	No	No	Yes
Group B (Size)	54	53	47
Rank	Yes	Yes	Yes
Pies	Yes	Yes	Yes
Differential	No	No	Yes
Likerts	Yes	Yes	Yes

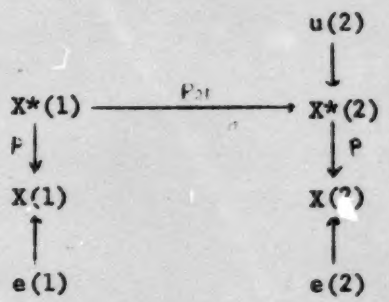


Figure 1. Path model for test-retest correlations adapted from Heise (1971).

Table 2. Test-retest correlations averaged across traits for evaluation scores and salience scores.

	Time 1-Time 2	Time 2-Time 3
	EVALUATION RELIABILITIES	
Group A		
Positivity	.597	.668
SDE	.704	.760
Group B		
Positivity	.577	.658
LE	.759	.782
Average Positivity	.587	.663
	SALIENCE RELIABILITIES	
Group A		
Indifferance	.486	.538
SFS	.576	.612
Ranks	.600	.656
Group B		
Indifference	.309	.554
LS	.653	.705
Ranks	.632	.598
Average Indifference	.401	.546
Average Rank	.615	.628

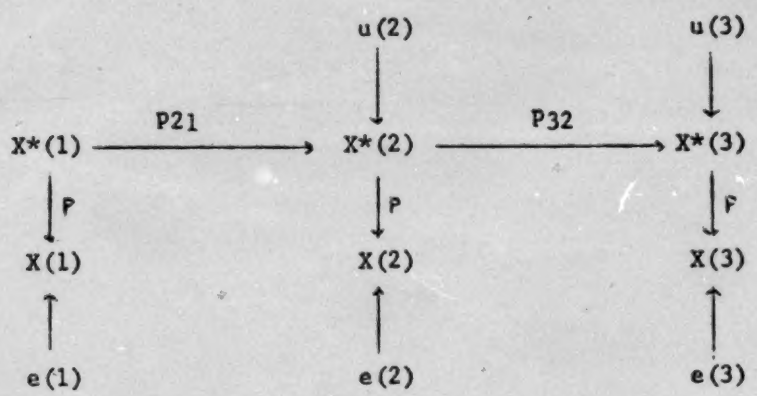


Figure 2. Path model for the 3-wave reliability model which separates true score stability from reliability (cf. Heise 1971).

Table 3. Reliabilities averaged across traits for the evaluation measures using Wiley and Wiley's constant error variance model.

	Time 1	Time 2	Time 3
Group A			
Positivity	.692	.704	.802
SDE	.753	.760	.860
Group B			
Positivity	.658	.677	.793
LE	.748	.852	.916

Table 4. Reliabilities averaged across traits for the salience measures using Wiley and Wiley's constant error variance model.

	Time 1	Time 2	Time 3
Group A			
Indifference	.555	.537	.727
SDS	.625	.572	.728
Ranks	.661	.551	.763
Group B			
Indifference	.591	.475	.720
LS	.705	.640	.787
Ranks	.621	.638	.775

Table 5. Validity diagonals for evaluative scores averaged across groups and times for pies, Semantic Differentials, and Likertscales.

	Pies vs SDE	Pies vs LE	SDE vs. LE
TRAITS (Rank Impt.)			
Parents (1)	.774	.582	.697
Blood(6)	.676	.691	.760
Inflation (5)	.216	.152	.688
Guns (8)	.648	.659	.804
Polkas (7)	.713	.688	.805
Parties (4)	.657	.659	.695
Radishes (9)	.667	.577	.852
Love (3)	.595	.679	.574
Summer (2)	.456	.485	.671
Alka-Seltzer (10)	.602	.608	.672
Average	.616	.592	.732

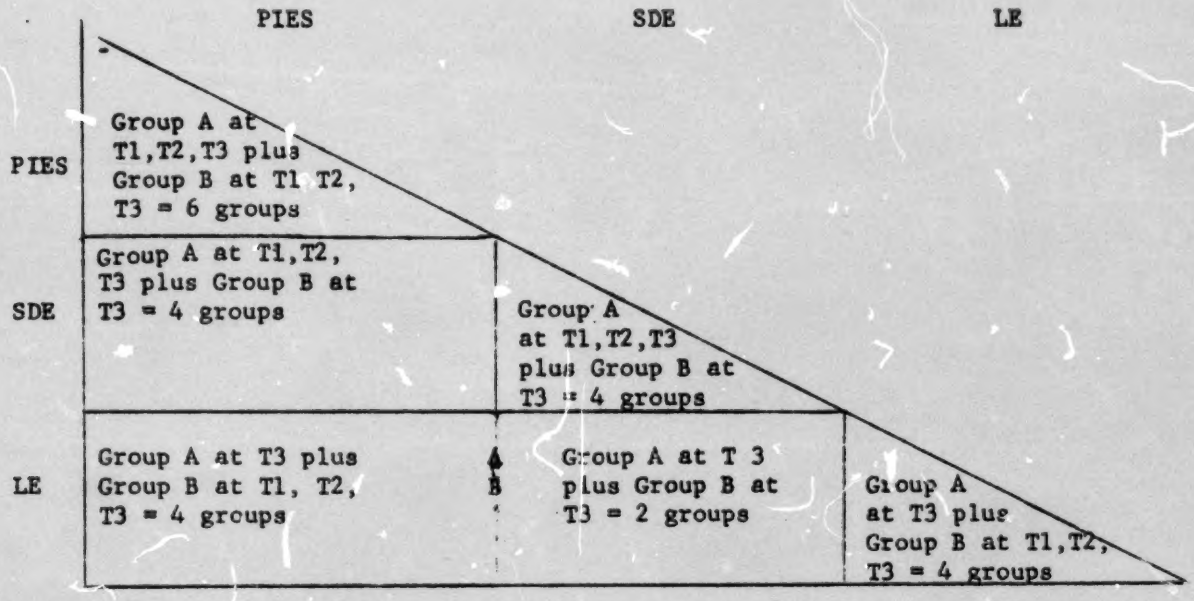

PIES
SDE
LE
PIES
Group A at T1,T2,T3 plus Group B at T1 T2, T3 = 6 groups
SDE
Group A at T1,T2, T3 plus Group B at T3 = 4 groups
Group A at T1,T2,T3 plus Group B at T3 = 4 groups
LE
Group A at T3 plus Group B at T1, T2, T3 = 4 groups
Group A at T 3 plus Group B at T3 = 2 groups
Group A at T3 plus Group B at T1,T2, T3 = 4 groups

Figure 3. MTMM matrix for evaluation scores averaged across groups and times as indicated in each block: Three methods by ten traits.

Table 6 Validity diagonals for salience scores averaged across groups and times for pies, SDS, LS, and Ranks.

	Pies vs SDS	PIES vs LS	SDS vs LS	PIES vs Ranks	SDS vs Ranks	LS vs Ranks
TRAITS						
Parents	.297	.214	.816	.187	.566	.624
Blood	.453	.297	.482	.184	.511	.441
Inflation	.338	.343	.551	.161	.164	.243
Guns	.091	.302	.180	-.032	.231	-.033
Polkas	.145	.292	.757	.290	.459	.436
Parties	.338	.344	.612	.368	.421	.428
Radishes	.002	.001	.611	.046	.095	.311
Love	.357	.027	.396	.337	.494	.276
Summer	.326	.291	.652	.205	.318	.384
Alka-seltzer	.044	.082	.332	.149	.332	.353
Average	.244	.222	.567	.192	.366	.356

Table 7. Mean salience scores for (1) Group A as determoned by pies, ranks, and SDS with corresponding numerical rankings and (2)Group B as determined by pies, rankings, and LS with corresponding numerical ranks.

	GROUP A			GROUP B		
	Pies	Ranks	SDS	Pies	Ranks	LS
TRAITS						
Parents	334.47(1)	2.19(1)	3.34(2)	338.73(1)	1.40(1)	2.69(1)
Blood	266.32(6)	5.44(6)	4.14(4)	271.57(6)	5.14(6)	7.41(8)
Inflation	274.3(5)	4.70(5)	4.47(5)	267.03(7)	4.86(5)	3.52 (2)
Guns	260.23(7)	8.20(9)	7.12(7)	287.43(5)	7.74(8)	6.65(6)
Polkas	184.10(8)	7.81(7)	8.07(8)	211.83(8)	7.51(7)	7.21(7)
Parties	287.97(4)	4.53(4)	5.88(6)	293.80(4)	4.68(4)	6.00(4)
Radishes	164.36(10)	8.34(10)	9.05(9)	186.97(9)	8.30(9)	10.45(9)
Love	325.26(2)	2.77(2)	3.27(1)	303.20(3)	3.52(3)	6.44(5)
Summer	315.03(3)	2.95(3)	3.96(3)	322.67(2)	3.19(2)	4.09(3)
Alka-Seltz	172.40(9)	8.17(8)	9.56(10)	177.77(10)	8.59(10)	12.07(10)

BIBLIOGRAPHY

Campbell, D.T. and D.W. Fiske
Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin,1959,56,81-105.

Cartwright,D. and F. Harary
Ambivalence and Indifference in generalizations of structural balance. Behavioral Science,1972, ??, 497-513.

Duncan, O.D.
Path Analysis: Sociological examples. IN H.M.Costner (Ed.), Sociological Methodology 1972. San Francisco: Jossey-Bass, 1972, 36-82.

Fishbein, M. and B. Raven.
An operational definition of belief and attitude. Human Relations, 1962,15, 35-44.

Hays, W.
Statistics for Psychologists. New York: Holt, Rinehart, & Winston, 1966.

Heise, D.R. Problems in path analysis and causal inference. In E.F. Borgatta (Ed.), Sociological Methodology 1969. San Francisco:Jossey-Bass,1968,38-73.

Heise,D.R.
Separating reliability and stability in test-retest correlation. In H.M. Blalock (Ed.), Causal Models in the Social Sciences. Chicago: Aldine-Atherton, 1971,348-363.

Hunter,J.E.
Dynamic Sociometry. Unpublished manuscript: Dept. of Psychology, Michigan State University, East Lansing, 1974.

Hylton, C. and W. Lashbrook
Apathetic and neutral audiences: A computer simulation and validation. Speech Monographs, 1972,39,105-113.

Jaccard,J., J. Weber, and J. Lundmark
A multitrait-multimethod analysis of four attitude assessment procedures. J. of Experimental Social Psychology, 11,1975,149-155.

Lashbrook, W.B. and La shbrook, V.J.
The effects of message sidedness for apathetic and neutral encoders engaged in a counterattitudinal task. Paper presented to the Annual Convention of the Speech Communication Assn., Houston, 1975.

Lashbrook, W.J. and T. Bush
An empirical investigation of apathetic and neutral attitude response to manipulations of source credibility, use of evidence and language intensity. Paper presented to the Western Speech Communication ASSn., Albuquerque, 1973.

Lashbrook, W.J. and J.Sullivan
Apathetic and neutral audiences: More on Simulation and validation. Speech Monographs, 1973,40, 317-321.

Lull, J.T. and P. Arnston
Attitude pie: A departure from linear measurement. Paper presented to the Annual Convention of the International Communication Assn.,New Orleans,1974.

Osgood,C.E., G.J. Suci, and P.H. Tannenbaum
The Measurement of Meaning. Urbana,Ill.; University of Illinois Press, 1957.

Sullivan, D. and W. Lashbrook
A simulation and validation of computer adaptations to apathetic audiences. Paper presented to the Western Speech Assn.,Newport Beach, 1974.

Torgerson, W.
Theory and Methods of Scaling. New York: Wiley, 1958.

Troyke,D. and W. Lashbrook
Computer modeling for the speech communication classroom. Paper presented to the Annual Convention of the Speech Communication Assn.,New York, 1973.

Wiley, D.E. and J.A. Wiley
The estimation of measurement error in panel data. In H.M. Blalock (Ed.), Causal Models in the Social Sciemces. Chicago: Aldine, 1971,364-373.

Wiley, J.A. and M.G. Wiley
A note on correlated errors in repeated measurements. Sociological Methods and Research, 3, 1974, 172-188.

APPENDIX A

Attitude Questionnaire

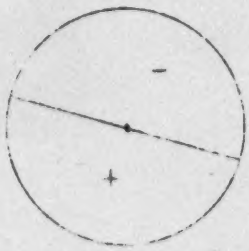
Please use the circles or "pies" on Page 3 to show what you think about the following attitude objects or concepts. Consider each circle as representative of the <u>total or sum of all your feelings</u> about the object or concept with which it corresponds. Think about each item separately. It may be that a certain part of you reacts <u>positively</u> to the object or concept. At the same time there may be another part of you that reacts <u>negatively</u> to it. Indicate these separate feelings by drawing in "slices" on the pies to reflect the amount of both your positive and negative reactions. For example, consider this attitude object:

"Economy Cars"

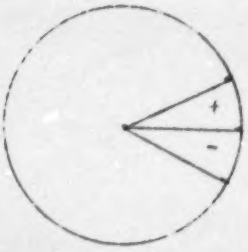
In many ways you may react positively to this object. Then again there may be some ways in which this object conjurs up a negative reaction in your mind. In this case, then, you could respond to the object by drawing in slices on the pie so that they look something like this:

Notice that you don't have to fill in the whole "pie." Put a plus sign (+) in the slice you draw to represent the degree of your positive reaction to the object or concept. Put a minus sign (-) in the slice which shows your negative reaction. If these two slices do not completely fill the pie (as in the case above), simply leave the remaining slice unmarked.

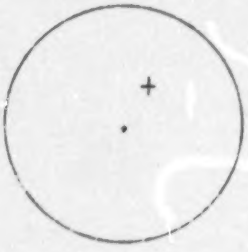
In this way you can use the pies to show the <u>strength</u> of your reaction as well as the relative amount of positive and negative feelings you have. Let's say, for example, that you react very strongly to an object or concept and have nearly equal positive and negative feelings about it. You could show this reaction by filling in the pie something like this:



On the other hand, if you have neutral feelings about the object or concept but only a weak general reaction to it, you can divide up the pie to look something like this:



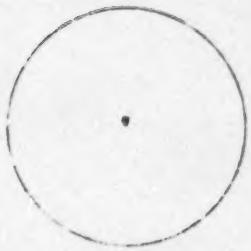
If your reaction to any object or concept is completely positive or completely negative, you can show this by putting in a plus or minus sign in the entire circle or any portion of it, depending upon the strength of your overall reaction:



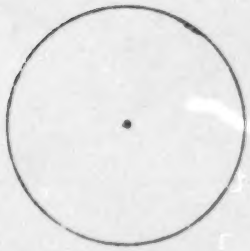
You have complete freedom to show precisely how you react to the attitude objects and concepts on the next page. Be sure to consider all your feelings about each item. Remember, the size of the slices you draw represents the strength of your positive and negative reactions.

Turn the page and begin.

Parents

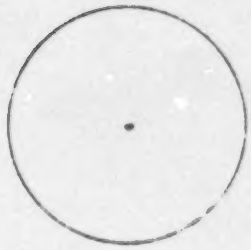


Parties



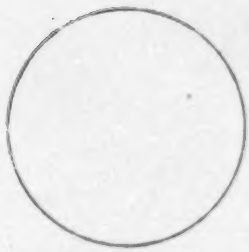
Giving Blood

Radishes

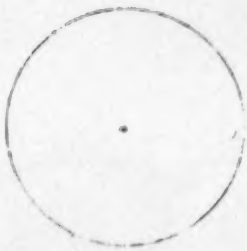


Inflation

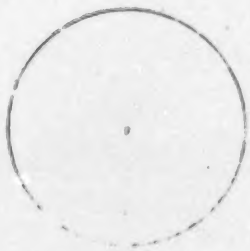
Making Love



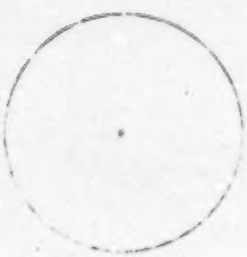
Handguns



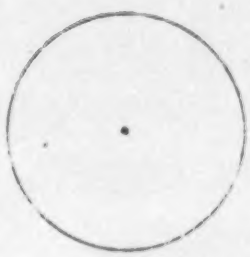
Summertime



Polkas



Alka-Seltzer



APPENDIX B

Note that four of the Semantic Differential scales are identical across traits. The fifth antonym pair peculiar to each trait is reported below:

Giving Blood

THE CHOICE OF WHETHER OR NOT TO GIVE BLOOD MEANS A LOT TO ME.

EVERYONE SHOULD DONATE BLOOD.

I DON'T WANT TO GIVE BLOOD.

THE IDEA OF GIVING BLOOD IS ON MY MIND A LOT.

THE THOUGHT OF GIVING BLOOD MAKES ME SICK.

Semantic Differential: pleasant-unpleasant

Polkas

WHEN I DANCE, I HATE TO POLKA.

MORE PEOPLE OUGHT TO LEARN TO POLKA.

POLKAS SHOULD NOT BE IGNORED.

WE OUGHT TO GET RID OF POLKAS.

POLKAS DON'T MEAN A THING TO ME.

Semantic Differential: worthless-valuable

Handguns

I WOULD LIKE TO OWN A HANDGUN SOMEDAY.

THE HANDGUN CONTROVERSY SHOULD BE A MORE IMPORTANT NATIONAL ISSUE.

HANDGUNS ARE HUMANE.

I NEVER THINK ABOUT HANDGUNS.

HANDGUNS SHOULD BE MADE ILLEGAL.

Semantic Differential: cruel-kind.

Making Love

THERE ARE SOME THINGS ABOUT THE THOUGHT OF MAKING LOVE THAT ARE DISGUSTING.

PEOPLE DON'T MAKE LOVE ENOUGH.

SOMETIMES THE THOUGHT OF MAKING LOVE SEEMS MORE BAD THAN GOOD.

THE IDEA OF MAKING LOVE IS NO MORE MEANINGFUL TO ME THAN ANY OTHER HUMAN BEHAVIOR.

I CAN GO FOR DAYS WITHOUT THINKING ABOUT MAKING LOVE WITH SOMEONE.

Semantic Differential: awful-nice

Summertime

SUMMERTIME IS PURE JOY.

I DON'T REALLY CARE ABOUT SUMMERTIME.

THERE IS NOTHING BETTER THAN SUMMERTIME.

I LIKE TO THINK ABOUT SUMMERTIME A LOT.

THERE'S A LOT OF THINGS ABOUT SUMMER THAT AREN'T SO GREAT.

Semantic Differential: sad-happy

Alka-Seltzer

I ALMOST NEVER THINK ABOUT ALKA-SELTZER

I CAN ONLY THINK OF GOOD THINGS TO SAY ABOUT ALKA-SELTZER.

JUST THE THOUGHT OF ALKA-SELTZER IS AWFUL.

IF I DRANK ALKA-SELTZER IT WOULD PROBABLY MAKE ME SICK.

ALKA-SELTZER MEANS A LOT TO ME.

Semantic Differential: nice-awful

Parties

I ALMOST ALWAYS ENJOY MYSELF AT A PARTY.

I DON'T CARE ABOUT PARTIES IN THE SLIGHTEST.

I WANT TO GO TO LOTS OF PARTIES.

I THINK ABOUT PARTIES MORE OFTEN.

PARTIES ARE A DRAG.

Semantic Differential: unpleasant-pleasant

Inflation

THERE'S NO NEED TO THINK ABOUT INFLATION.

IT'S ALRIGHT IF THE RATE OF INFLATION KEEPS CLIMBING.

INFLATION IS EVERYBODY'S BUSINESS.

A NATION IS IN BAD SHAPE WHEN INFLATION HITS ITS ECONOMY.

PEOPLE SHOULD NOT BE PLEASED WITH INFLATION.

Semantic Differential: pleasant-unpleasant

Parents

MY PARENTS ARE OFTEN INSENSITIVE TO THE NEEDS OF OTHERS.

MY PARENTS ARE EXTREMELY IMPORTANT TO ME.

MY PARENTS ARE VERY GOOD PEOPLE.

THERE ARE REALLY ONLY A FEW POSITIVE THINGS TO SAY ABOUT MY PARENTS.

MY PARENTS REALLY DON'T MATTER TO ME.

Semantic Differential: kind-cruel

Radishes

RADISHES TASTE TERRIBLE.

WE SHOULD THINK MORE ABOUT RADISHES.

RADISHES MEAN ALMOST NOTHING TO ME.

RADISHES ARE A PLEASANT ADDITION TO A MEAL.

THERE'S NOTHING WRONG WITH RADISHES.

Semantic Differential: pleasant-unpleasant